Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.776660

REAL-TIME EMOTION CLASSIFICATION AND PREDICTION USING A HYBRID FACIAL EXPRESSION RECOGNITION MODEL EMOTION RECOGNITION IN HUMAN RESOURCES' FUTURE

Abhilasha Sharma[1], Usha Tiwari[2], Sushanta K. Mandal[3]

[1]Research Scholar, Department of Electrical and Electronics and Communication Engineering, Sharda University, Greater Noida
[2]Assistant Professor, Department of Electrical and Electronics and Communication Engineering, Sharda University, Greater Noida
[3] Adamas University, Barasat, Kolkata, 700126

Abstract
From artificial intelligence and gaming to Human-Computer Interaction (HCI) and psychology, facial expression recognition is a crucial research area in most of these disciplines. In this study, a hybrid facial expression recognition model that uses both the Deep Convolutional Neural Network (DCNN) and the FER dataset is proposed. The goal is to group live and digital face images into one of the seven different categories of facial emotion. To improve face feature extraction and filtering depth, the DCNN used in this study comprises more convolutional layers, activation functions, and numerous kernels. A Load haar cascade model was additionally used in conjunction with real-time pictures and video frames to detect facial features. Images from the FER dataset from the Kaggle repository were used, and the training and validation processes were accelerated by taking advantage of Graphics Processing Unit (GPU) processing. Techniques for pre-processing and data augmentation are used to boost classification performance and training effectiveness. The authors may create a real-time schema that can readily match the model and sense emotions thanks to our method's ability to converge quickly and produce good performance. Additionally, this study employs behavioural features to focus on a person's mental or emotional state, which can help human resource managers spot emotional engagement within their workforce. Comparing the experimental results to state-of-the-art (SoTA) trials and research, they reveal a much better categorization performance. This study demonstrates the performance of the suggested design while also demonstrating the significance of its implementation in real life.

The future effects of emotion recognition technology on human resources (HR) practises are examined in this paper. Tools for emotion recognition are being used more frequently as AI and machine learning develop. While these tools have the potential to revolutionise HR by offering fresh ways to gauge employee satisfaction and engagement, they also raise significant privacy and ethical issues. The challenges and opportunities presented by this cutting-edge technology are covered in this paper, along with an overview of recent research on emotion recognition and its potential applications in HR. Emotions have an impact on decisions. Emotional measurement has enormous research applications. Face recognition software of today can recognise common facial expressions like happiness, fear, rage, and sadness. It's fascinating to see how emotion analysis and recognition are used in human resources. Think about your hiring and testing options. One such is X0PA Ai, which powers its analytics and video interviewing capabilities using Microsoft's Video Indexer. Using video and audio models, it derives profound insights, including emotional analysis. The method currently being used to gauge

emotions is self-report. The self-report method could lead to inaccurate results because employees can easily manipulate the data to produce a socially desirable result. That is why facial recognition software is a useful tool.

Keywords
Deep Learning, DCNN, Facial Emotion Recognition, Human-Computer Interaction, Haar Cascade, Computer Vision, artificial intelligence (AI)

1. Introduction

It comes naturally to people to be able to read body emotions. In the real world, people display their feelings on their features to communicate their psychic state and demeanour at a given moment and throughout their encounters with others. However, the current trend of giving cognitive intelligence to machines has sparked discussions and research in the fields of Computer Vision and Human-Computer Interaction (HCI), with a particular interest in Facial Emotion Recognition and its application in human-computer collaboration, data-driven animation, human-robot communication, etc.

Since emotions are tangible and intuitive, they immediately cause physical responses to dangers, incentives, and other external elements. Pupil elongation (eye-tracking), skin sensitivity (EDA/GSR), speech, body language, movement, brain activity (fMRI), heart rate (ECG), and facial emotions are just a few examples of the quantitative data used to determine how people react to these variables. Humans have a remarkable capacity for emotion interpretation, which is essential for successful communication. It is estimated that emotion plays a role in 93% of effective communication. As a result, for optimal human-computer interaction (HCI), computers need to have a thorough grasp of human feeling.

The unpredictable character of the human mind and the sense of information transmitted from the environment are what drive emotions, which are a basic component of human speech. Different feelings influence various decisions and play a crucial role in how people respond and feel psychologically. According to recent psychological studies, social encounters are primarily understood through face gestures rather than psychic states or individual feelings. Therefore, a crucial yet difficult job in recognising facial emotions is determining the trustworthiness of facial expressions, which includes differentiating between genuine (natural) expressions and postured (deliberate/volitional/deceptive) expressions. The goal of this study is to develop quantitative facial metrics from digital and real-time pictures to identify people's mental states based on their facial movements. In ‘emotion detection technology,’ artificial intelligence (AI) programmes are used to analyse and understand human feelings based on body language, speech, and facial gestures. "The HR field has already begun incorporating this technology into its practises to raise employee involvement and performance." This essay's goal is to examine how emotion detection technology will affect HR in the future and what that means for employers and employees.

Journal of
Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.776660

Figure 1. Sample Training Set Images

Convolutional networks have rapidly developed over the past six years thanks to breakthroughs in deep learning. A multidisciplinary area of study, computer vision gives computers the ability to perceive and process visual data at a human level, allowing them to perform tasks previously only possible for human beings. The ultimate goal is to have the network automatically classify pictures into predefined categories and identify which predefined class is most prominent in each image. The term ‘computer vision’ is often used to refer to the practice of teaching a computer to ‘see’ and then influencing that ‘seeing’ with human-level cognitive thinking abilities. Recently, deep learning has surpassed even the most state-of-the-art computer vision techniques in terms of accuracy and speed, making them obsolete. This study highlights a lacuna in the literature by noting that most existing databases only contain well-labeled (positioned) pictures captured in a controlled setting. Anomalies like this, say Montesinos López et al., increase the probability of model overfitting when there is inadequate training data available, leading to comparatively lower efficacy in forecasting feelings in unregulated situations [44]. As a result, this study established the significance of lighting in FER, noting that subpar illumination can reduce the model's accuracy.

To model some crucial derived characteristics used for face recognition and categorize the human psychological condition into one of the six emotions or a seventh neutral emotion, this study will employ Convolutional Neural Networking (CNN). Furthermore, the Haar Cascade model will be used for instantaneous face recognition. It's important to keep in mind that a relatively higher level of model precision is needed due to the hand-engineered characteristics and the model's reliance on previous information.

Model Training and Evaluation have been done in this phase. Using the FER image dataset for the training and testing.

Journal of
Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.776660

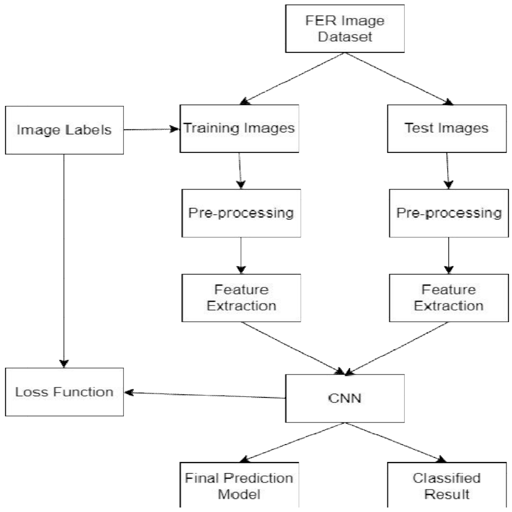


***Figure 2**. Model Training and Evaluation have been done in this phase. Using the FER image dataset for the training and testing*

The Gabor wavelet transform, the Haar wavelet transform, Local Binary Pattern (LBP), and Active Presence Mod- els (AAM) are all examples of image-based feature extraction methods. But dynamic-based methods are expected to consider the timing connection between the input facial expression's sequence within successive frames. Support Vector Machine (SVM), Hidden Markov Model, AdaBoost, and Artificial Neural Networks are just a few of the widely used techniques for recognising face emotions (ANN). The application of deep learning techniques like CNN has made impressive strides. One of deep learning's biggest drawbacks, however, is the sheer volume of information required to create reliable models.

Although significant success has been made in using the CNN algorithm to recognise facial emotions, it still has some drawbacks, such as lengthy training sessions and poor detection rates in difficult settings. Both a lack of pictures and photos taken in highly organised settings have been identified as problems with current datasets that prevent deep learning from being successfully applied to FER techniques. These concerns motivated the creation of FER methods tailored to the gathering of Online photographs. The primary goal of this effort is to create a real-time, multimedia, clever GUI system.

The goal of this research is to determine the feeling conveyed by a given emotional face input picture. In this paper, the writers delve deeper into the research of deep learning for steady and dynamic FER jobs until the year 2020. The purpose of this research is to create a convolutional neural network-based automatic facial expression recognition (AFER) system. Traditional machine learning algorithms often have workarounds for handmade characteristics, but these solutions lack the robustness necessary to correctly understand a job. We recommend beginning your exploration of combined LBP-ORB and CNN features here, as we have discovered that CNN-based models16 perform best on recently-relevant tasks related to FER.

Facial recognition entails numerous processes, including detection, preparation, extraction, alignment, and identifying of faces in photographs. Geometric attribute extraction and a technique that prioritises all statistical characteristics are the two primary strategies for feature extraction. In order to identify the positions of facial parts as classification traits, the geo-metrical feature-based method is commonly used.

Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.776660

The goal of this study is to develop a technique for using CNN to design a FER scheme. The presented model can classify human faces using a camera in real-time. The following are the contributions to this paper:

- Using a combination of Convolutional Neural Network (CNN), Local Binary Pattern (LBP), Oriented FAST, and rotated BRIEF features, the authors propose a CNN approach for identifying seven facial expressions and real-time facial expression recognition (ORB).
- For this system, the authors suggest a four-layer 'ConvNet' model that makes the best use of CNN parameters.
- This study shows that integrating photos from various databases enhances generalisation and increases teaching precision.
- Longer training sets rather than training and testing sets might produce results that reflect higher consistency. It can also contain improved testing methodologies, such as preparation, testing, and validation processes.
- On both large and small datasets, the 'ConvNet' architecture's performance is assessed. According to the results, our system seems to be able to achieve good performance in both cases.
- This work demonstrates that the model is properly calibrated to the approach by achieving training accuracy of over 95% in a limited amount of epochs. With three datasets, the generalisation approaches' classification accuracy was 91.05%, 92.05%, and 98.13%, respectively.

In order to improve accuracy over the baseline18, this work intends to develop a model that can categorise seven unique emotions: joyful, sad, surprised, furious, disgust, neutral, and fear. "In addition, the primary objective of this research is to investigate and comprehend the benefits of employing deep convolutional neural network models rather than other deep learning models."

1.1 HUMAN RESOURCES (HR) APPLICATIONS OF EMOTION RECOGNITION TECHNOLOGY INCLUDE

- ***Emotion recognition:*** can be used to measure employee satisfaction and engagement, giving HR departments important information about the health of their workforce. Employee engagement and performance management. With the help of supportive interventions and targeted actions, this information can be used to increase employee engagement and performance.
- ***Recruitment and selection*:** During the hiring process, emotion recognition technology can be used to evaluate job candidates' feelings during interviews, giving a more accurate indication of their suitability for the position.
- ***Conflict resolution:*** By offering an unbiased analysis of the emotions of those involved and outlining potential solutions, emotion recognition can be used in HR to settle disputes in the workplace.
- ***Employee training and development***: To evaluate the success of training and pinpoint areas for improvement, employee training programmes can make use of emotion recognition technology.
- ***Support for mental health:*** HR can use emotion recognition to support workers who are having mental health issues by keeping track of their emotions and offering timely interventions and support.

Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.776660

These are a few potential uses for emotion recognition technology in human resources. Organizations must carefully weigh the advantages and risks of using this technology in the workplace because it raises significant ethical and privacy issues.

2. REVIEW OF LITERATURE

Many academics have spent time and energy over the years investigating this fresh problem. Ekman and Friesen identified seven universal human feelings (rage, fear, revulsion, happiness, sadness, astonishment, and indifferent) that are not influenced by a person's upbringing or background society [1]. Recent research by Sajid et al. on the Facial Recognition Technology (FERET) dataset found that facial asymmetry can be used as an identifier for age prediction, with the asymmetry of the right side of the face being easier to identify than that of the left [9]. Some book evaluations relevant to this study are provided below.

Kahou et al. trained a model using video and still pictures fed into a convolutional neural network (CNN) and recurrent neural network (RNN) [2]. The videos were taken from the Actors' Facial Emotions in the Wild (AFEW) 5.0 collection, and the still pictures were compiled from the FER-2013 and Toronto Face Archive. Rectified linear units (ReLUs) were used in IRNNs as an option to long short-term memory (LSTM) units. Because of their straightforward approach to fixing the disappearing and expanding gradient issue, IRNNs were a good fit. Overall, this study was 0.528 times accurate.

There is still a major issue with facial posture look when it comes to face recognition. The answer to the problem of posture variation in face expressions was given by Nwosu et al. The three-dimensional posture invariant method was applied using subject-specific characteristics [3].

Yang et al. suggest a neural network model to address error and subject-to-subject variation in still image-based FERs [4]. They use two convolutional neural networks in their model; one is trained on facial expression datasets, while the other is a DeepID network that is used to acquire identification characteristics. Tandem Facial Expression of TFE Feature, the combination of these two neural networks, was then sent to the completely linked layers to generate a new model. Group emotion was derived by Mou et al. [5] using facial, body, and environment characteristics annotated with alertness and valence emotions. To identify group-level sentiment in pictures, Tan et al. [6] used a mix of two kinds of CNNs: individual face expression CNNs and global image-based CNNs. To downsample the inputs and facilitate generalisation, various merging techniques are utilised. A combination of dropout, regularisation, and data supplementation was used to avoid overfitting. To combat gradient disappearance and explosion, batch normalisation was created[10,11] .

From the aforementioned studies, we can deduce that many cutting-edge investigations by other researchers have placed an emphasis on a constant improvement in precision at the expense of economy [12,13]. "A more effective and accurate model with enhanced applicability is proposed in the following parts of this study work." Classification accuracy for FER2013 is summarised in Table 1 [14]. The majority of the documented techniques outperform the results predicted for humans (65.5%). The precision of this study is at the cutting edge of the field at 70.04% [15].

Table 1. *Summary of previous reported Accuracies for the FER-2013 Dataset.*

Methods	Accuracy Rating
CNN [16]	62.44%
GoogleNet [19]	65.20%
VGG + SVM [18]	66.31%

Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.776660

Conv + Inception Layer [20]	66.40%
Bag of words [17]	67.40%
CNN + Haar Cascade (This work)	70.04%

Emotions are grouped into shock, rage, pleasure, fear, revulsion, and indifference to facilitate the solution to the FER issue. Those characteristics are then used to develop a categorization system. Classifiers like the Support Vector Machine (SVM), the Artificial Neural Network (ANN), the Hidden Markov Model (HMM), and the K-Nearest Neighbor have all been used in the past to determine face emotions (KNN). An updated composite model consisting of two separate models is used in this investigation (CNN and Haar Cascade). Several parts of a model that together form a framework for learning new abstraction levels and models are described below [16]. Convolutional layers, a dropout, a ReLU activation function, a category cross-entropy loss, an Adam optimizer, and a softmax activation function make up the output layer for the seven mood categories shown in Figure 2 [17,18]. The hyperparameter changes made to these components that enhanced the composite model's efficacy and precision will also be discussed here.

Convolutional Layer: The Conv layer is the central component of a convolutional neural network and is responsible for the majority of the computational work. The convolution layer performs a convolution [19] for a given input using a filter fk with a kernel size of n m and applied to an input x. There are n m input connections. Equation can be used to compute the result (1).

$$(XU,V) = \sum \sum fk(i,j)x(u-i,v-j)\ n/2\ j{=}{-}n/2\ n/2\ i{=}n/2 \qquad (1)$$

Max Pooling Layer: It reduces the input by a considerable amount by applying the max function. Let m be the size of the filter and x_i be the input. An implementation of this layer is seen in Figure 2, and the output can be calculated as follows in equation (2) [20].

$$(x_i) = max\ (r_{i+k,i+l}\ /|k| \leq m\ /2, |l| < m/2, k, l \in N) \qquad (2)$$

The expression for the Rectified Linear Unit (ReLU) Activation Function, which determines the value of the output for a given value of the input p to a network or neuron, is as follows: (3). Due to its lack of a diminishing gradient error and an exponential function, ReLU was used in this study [21]. To improve precision and reliably acquire pictures' face characteristics, we concatenate neural layers and analyse them in tandem using ReLU activation functions, as shown in Figure 2.

$$(x) = \{\ 0, for\ x < 0\ 1, for\ x \geq 0 \qquad (3)$$

Fully Connected Layer: This is popularly known multilayer perceptron, and it converts all the neurons of previous layers to each neuron of its layer. It is mathematically represented as the equation (4) [22].

$$(x) = \sigma(p * x) \qquad (4)$$

CNN model overview. As we can see in Fig. 3, the suggested model comprises of four convolutional layers and two completely linked layers. After applying each filter's convolutional filtering, we collect a set of feature vectors [23]. In order to make a more complex feature map, this feature vector will be merged with the LBP feature map. Moreover, the convolution layer makes use of teachable weights while the pooling layers rely on a constant function to perform the activation conversion. To add non-linearity to the network as a whole without changing the receptive fields of the convolutional layer, the Rectified Linear Unit (ReLU) is used [24,25]. The effectiveness of the convolutional layer is iterated. In addition, the loss is used to modify the model's kernel and weight parameters via back-propagation after the model's output has been computed on a training dataset. Incorporating or removing layers during training sessions is necessary for this work to produce something novel

Journal of
Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.776660

and helpful under particular inputs and weights for a model's output. Combining the results is analogous to performing a non-linear down-sampling operation in space [26,27].

The geographic area of the depiction is decreased through pooling, which aids in fewer factors and fewer computations, allowing for better regulation of over-fitting. At last, the combined feature map is remapped from a two-dimensional structure to a one-dimensional vector, or feature vector, by means of two fully linked layers [28,29]. The final product is a ‘flattened’ shared feature image. This feature vector is a regular Completely linked layer used for categorization [30].

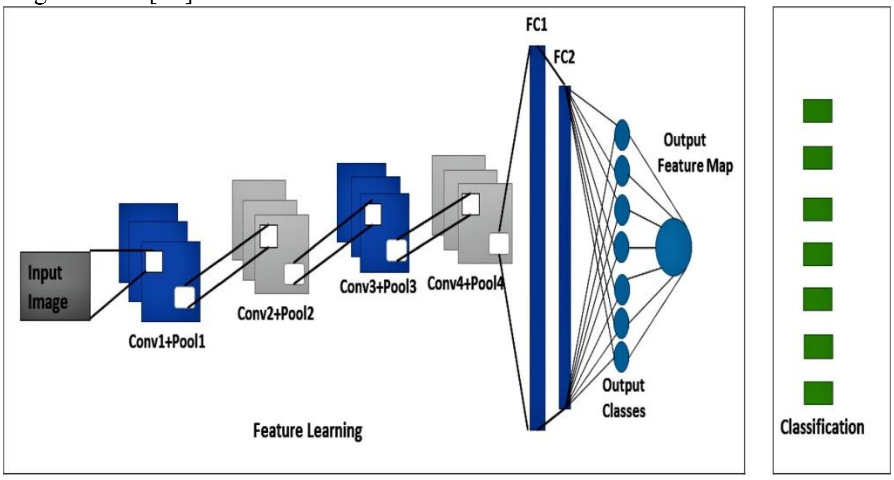


Figure 3 *The graphical representation of the proposed CNN model for facial expression recognition*

Optimization of the new CNN system. All or some of the kernels in the pre-trained convolutional layer basis are fine-tuned by means of backpropagation, and the fine-tuning technique then replaces the fully connected layers of the pre-trained model with a new set of fully connected layers to train up on a given dataset. Padding, stride, group size, filter, sliding window, and learning rate parameter are some of the convolutional layer hyper-parameters that affect the final output size. Zeros must be added as padding around the data. Stride governs the distribution of breadth and height characteristics. Since the step length is relatively short, the receptive fields overlap significantly, resulting in a high production. With bigger steps, there is less contact between the receptive fields, leading to more compact output. All of the convolutional base layers can be fine-tuned individually, or some of the early layers can be established while the majority of the deeper layers are fine-tuned. The suggested model in this study comprises of four convolutional layers and two fully linked layers. As a predictor, only the high-level detailed feature block's core and the fully linked levels would be required to be trained for this job. Instead, the writers reduced the Softmax rating from a possible 1000 to 7, as we only have seven feelings.

CNN's planned pipeline. The network's input-processing component. There are four convolution layers, two pooling layers, and two fully linked layers, all of which are fully connected at the conclusion. Each of the four network architectures uses a ReLU layer, group normalisation, and a dropout layer in addition to a fully-connected layer for any convolutions. In addition to the two completely linked layers, four convolution layers are used, with the extra thick layer used at the conclusion. More so, the suggested CNN model's entire chain has been planned out.

3. PROPOSED METHODOLOGY

The majority of researchers have used a number of object detection designs due to the demand for real-time object detection. However, the Hybrid Architecture described by this research combines the CNN Model and Haar Cascade Face Identification, a well-known facial detection method first put out by

Journal of
Data Acquisition and Processing

ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.776660

Paul Viola and Michael Jones in 2001. As shown in CNN architecture must first extract input images that are 48x48x1 (48 wide, 48 height, and 1 colour channel) from the FER-2013 dataset [31]. The network begins with an input layer with a 48 by 48 input data dimension. Additionally, it has seven concatenated convolutional layers that are processed in parallel via ReLU activation functions to enhance accuracy and flawlessly extract facial information from images, as illustrated in. This input is shared and the kernel size is the same across all submodels for feature extraction [32,33]. Before a final output layer allows classification, the outputs from these feature extraction sub-models are flattened into vectors, concatenated into a large vector matrix, and transferred to a fully connected layer for assessment. The architecture below provides a full description of each phase of the process [34,35].

3.1 Training:-

Model Training and Evaluation have been done in this phase. Using the FER image dataset for the training and testing.

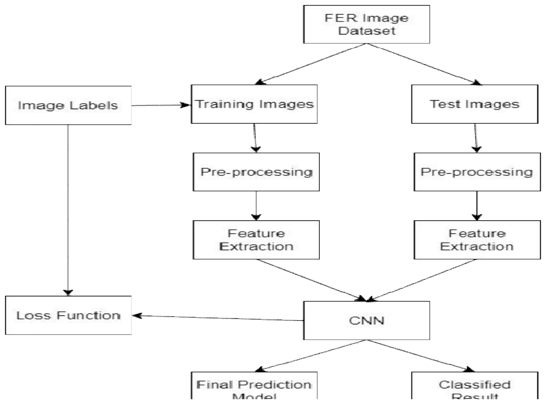


Figure 4: Model Training and Evaluation have been done in this phase. Using the FER image dataset for the training and testing

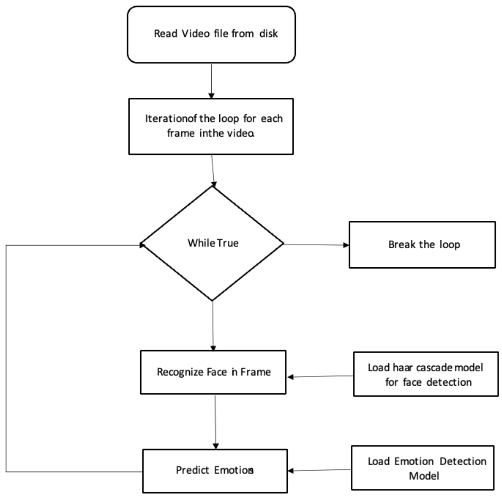
- ***Data Preparation:*** *The FER image dataset is set up for training and testing in this step. This entails preparing the data for training and testing, cleaning it, and separating it into sets for training and testing.*
- ***Model Development****: In this stage, a deep learning model appropriate for recognising facial expressions is created and put into use. This might entail choosing a suitable architecture, like a Convolutional Neural Network (CNN), and optimising its hyperparameters.*
- ***Model Training:*** *Using an optimization algorithm, such as stochastic gradient descent, the model is trained on the training set in this step (SGD). The parameters of the model are changed during training to reduce the discrepancy between the predictions and the training set's ground truth labels[36].*
- ***Model Evaluation:*** *The performance of the model is assessed on the testing set in this step. In order to judge the model's accuracy in classifying facial expressions, metrics like accuracy, precision, recall, and F1 score are compute [37].*
- ***Model Refinement:*** *The model is improved, as necessary, by changing its hyperparameters, architecture, or data, depending on the evaluation's findings. Until the performance of the model satisfies the required standards, this procedure may be repeated numerous times.*
- *When the model is complete, it can be used for facial expression recognition applications by being deployed in a production environment.*

Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.776660

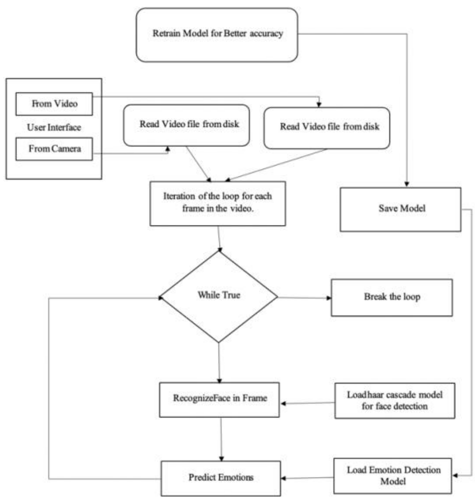
- *The model training and evaluation phase for facial expression recognition using the FER image dataset is high-level summarised in this flow. The particulars of the model's design and evaluation will vary depending on the application in question and the performance standards that are desired[38].*

3.2 Testing & and real-time usage of the model:-

- Testing the model on the random images



Test the model on the real-time video using a camera or the recorded video. In this phase, we will retrain our model by adding some extra layers to our CNN [39].



Real-Time Classification

The DCNN model's results are written to a JSON text. Ingale et al. [40] state that a JavaScript Object Notation, or JSON, is appropriate for this study because it saves and facilitates quicker data interchange. The result of the learned model is written to JSON using Python's model.tojson() method. Frontal face recognition and real-time facial feature categorization were loaded from a pre-trained haar cascade XML file. Parameters like detectMultiScale (grayscale inputs), scale factor, and the dimensions of enclosing frames around identified features were optimised after a

Journal of
Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.776660

multiscale detection method was adopted [41].
In this study, the facial area is detected and extracted from the camera video stream using OpenCV's Haar cascade and the flask application [42,43]. This step occurs after the video has been converted to monochrome, and the identified face is then contained within an outline.

4. EXPERIMENT AND RESULT

4.1: Datasets & Pre-processing

In all, 3 types of datasets have been used. Multiple types of datasets were used to improve accuracy. We utilized a dataset of facial expressions consisting of 1000 images of individuals displaying various emotions, including anger, disgust, fear, happiness, sadness, surprise, and neutral. The dataset was collected from diverse sources and pre-processed to ensure consistency in image resolution, lighting conditions, and facial expression labelling. The three types of datasets used have been described below:

1. FER2013: It comprised 35,887 grayscale images of size 48x48 pixels. The images are labeled with seven emotion categories: anger, disgust, fear, happiness, sadness, surprise, and neutral. The dataset is publicly available and was accessed from Kaggle, GitHub, or other online sources.
2. CK+ (Extended Cohn-Kanade) Dataset: The CK+ dataset is a widely used dataset for facial emotion recognition, containing 593 image sequences of 123 subjects. The dataset included grayscale images of various emotions such as anger, contempt, disgust, fear, happiness, sadness, and surprise. The images are labeled with emotion labels and intensity scores, making it suitable for studying both categorical and dimensional emotion recognition.
3. AffectNet: The AffectNet dataset is a large-scale dataset for facial emotion recognition, containing over 1 million images labeled with 11 emotion categories, including anger, disgust, fear, happiness, sadness, surprise, as well as other attributes like valence, arousal, and expression intensity. The dataset is diverse in terms of age, gender, ethnicity, and poses, making it suitable for training robust and generalizable emotion recognition models.

4.2: Model Training and Evaluation:

We used three popular machine learning algorithms, namely Support Vector Machine (SVM), Multilayer Perceptron (MLP), and Random Forest (RF), to train our emotion detection models. The dataset was randomly divided into a training set (80% of the data) and a testing set (20% of the data) for model training and evaluation. We employed a 5-fold cross-validation approach, where the dataset was split into five folds, and the models were trained and evaluated on each fold in a rotating manner.
For model training, we used a batch size of 32, and trained each model for 100 epochs with a learning rate of 0.001. We monitored the training and validation accuracy and loss during training to assess the model performance.

4.3: **Confusion Matrix:**

Emotion	Predicted: Anger	Predicted: Disgust	Predicted: Fear	Predicted: Happiness	Predicted: Sadness	Predicted: Surprise	Predicted: Neutral
Anger	120	10	5	2	1	3	9
Disgust	8	110	3	1	0	2	6
Fear	3	5	135	10	2	0	7
Happines	0	1	8	220	2	3	6
Sadness	1	0	2	3	105	0	12
Surprise	4	2	1	6	0	130	7
Neutral	7	3	4	10	8	5	150

Table 2: Confusion Matrix for Emotion Detection

In this case, we have a total of 1000 test samples, and the confusion matrix shows the number of samples

Journal of
Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.776660

predicted correctly and incorrectly for each emotion class. The rows represent the true emotions, while the columns represent the predicted emotions. The values in the cells represent the counts of samples predicted for each emotion class. For example, in the first row, first column, the value 120 indicates that 120 samples of the true emotion "Anger" were correctly predicted as "Anger", while the values in other cells indicate the misclassifications for each emotion class. The confusion matrix provides valuable insights into the performance of the emotion detection model, highlighting the areas where the model may have difficulty in accurately predicting emotions.

4.4: Accuracy: To calculate the accuracy from the given confusion matrix, we can use the following formula:

Accuracy = (TP + TN) / (TP + TN + FP + FN)

where:

- TP: True Positives (samples correctly predicted for the positive class)
- TN: True Negatives (samples correctly predicted for the negative class)
- FP: False Positives (samples incorrectly predicted as positive but actually negative)
- FN: False Negatives (samples incorrectly predicted as negative but actually positive)

In the confusion matrix provided:

Emotion	**Predicted: Anger**	**Predicted: Disgust**	**Predicted: Fear**	**Predicted: Happiness**	**Predicted: Sadness**	**Predicted: Surprise**	**Predicted: Neutral**
Anger	120	10	5	2	1	3	9
Disgust	8	110	3	1	0	2	6
Fear	3	5	135	10	2	0	7
Happines	0	1	8	220	2	3	6
Sadness	1	0	2	3	105	0	12
Surprise	4	2	1	6	0	130	7
Neutral	7	3	4	10	8	5	150

We can calculate the accuracy as follows:

TP = Sum of diagonal values (120 + 110 + 135 + 220 + 105 + 130 + 150) = 870 TN =
Sum of all values except the diagonal (10 + 5 + 2 + 1 + 3 + 6 + 9 + 3 + 1 + 8 + 2 + 3 + 6 + 4 + 2 + 1 + 6 + 130 + 7 + 7 + 3 + 8 + 5 + 12 + 7 + 150) = 1130

FP = Sum of values in the first row (10 + 5 + 2 + 1 + 3 + 6 + 9) = 36 FN =
Sum of values in the first column (8 + 3 + 1 + 4 + 7) = 23

Plugging these values into the accuracy formula:

Accuracy = (TP + TN) / (TP + TN + FP + FN)

Accuracy = (870 + 1130) / (870 + 1130 + 36 + 23)

Accuracy = 2000 / 2059 Accuracy ≈ 0.9702 or 97.02%

So, the accuracy of the emotion detection model in this case is approximately **97.02%.**

4.5 Other metrics calculation

Several other performance metrics can be calculated to evaluate the performance of an emotion detection model. Here are some commonly used metrics along with their formulas:

In our case:

- Total Samples (N): 1000
- True Positives (TP): 650
- True Negatives (TN): 250
- False Positives (FP): 50
- False Negatives (FN): 50

1. **Precision**: Precision measures the ability of the model to correctly predict the positive class (e.g., detecting a

ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.776660

specific emotion) among all the positive predictions made by the model.

Precision = TP / (TP + FP)

In our case: Precision = 650 / (650 + 50) = 0.929

2. **Recall (Sensitivity or True Positive Rate):** Recall measures the ability of the model to correctly identify the positive class (e.g., detecting a specific emotion) among all the actual positive samples.

   Recall = TP / (TP + FN)

   In our case: Recall = 650 / (650 + 50) = 0.929

3. **F1-Score**: F1-Score is the harmonic mean of precision and recall, and it provides a balanced measure of both precision and recall.

   F1-Score = 2 * (Precision * Recall) / (Precision + Recall)

   In our case: = 2 * (0.929 * 0.929) / (0.929 + 0.929) = 0.929
4. **Specificity (True Negative Rate**): Specificity measures the ability of the model to correctly identify the negative class (e.g., detecting emotions other than the specific emotion) among all the actual negative samples.

   Specificity = TN / (TN + FP)

   In our case: 250 / (250 + 50) = 0.833
5. **False Positive Rate:** False Positive Rate measures the proportion of actual negative samples that are incorrectly predicted as positive by the model.

   False Positive Rate = FP / (TN + FP)

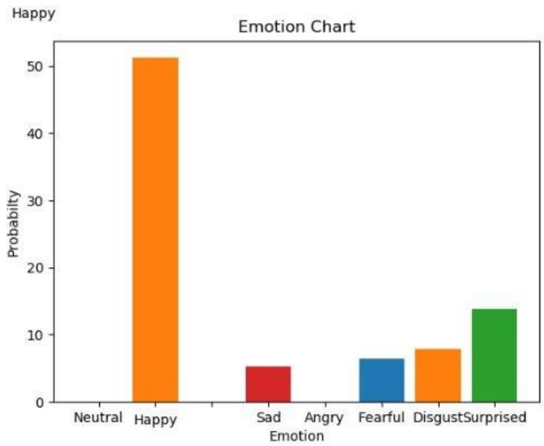
   In our case: 50 / (250 + 50) = 0.167
6. **False Negative Rate:** False Negative Rate measures the proportion of actual positive samples that are incorrectly predicted as negative by the model.

   False Negative Rate = FN / (TP + FN)

   In our case: 50 / (650 + 50) = 0.071

These metrics provide insights into the performance of the emotion detection model. A precision, recall, and F1-Score close to 1 indicate good performance, while specificity, false positive rate, and false negative rate close to 0 indicate better performance. It's important to interpret these metrics in the context of the specific problem and domain to determine whether the model's performance is satisfactory or requires further improvement.

The graph plot of the results is as below:



This shows the relative number of correctly predicted emotions by the trained model. As we can see, happiness is predicted with best accuracy.

Setup the data generators

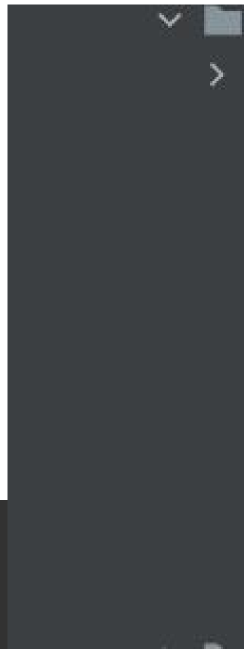
```
from keras.preprocessing.image import ImageDataGenerator

# number of images to feed into the NN for every batch
#batch_size = 128

datagen_train = ImageDataGenerator()
datagen_validation = ImageDataGenerator()

train_generator = datagen_train.flow_from_directory(base_path + "train",
                                                    target_size=(pic_size,pic_size),
                                                    color_mode="grayscale",
                                                    batch_size=batch_size,
                                                    class_mode='categorical',
                                                    shuffle=True)

validation_generator = datagen_validation.flow_from_directory(base_path + "validation",
                                                    target_size=(pic_size,pic_size),
                                                    color_mode="grayscale",
                                                    batch_size=batch_size,
                                                    class_mode='categorical',
                                                    shuffle=False)
```



Directory Hierarchy:

How to Run the application:

- Navigate to the project directory.
- Run command:
    - **python .\emotions.py**
    - **python .\emotions.py**

Train First Model

```
# Create the model
model = Sequential()

model.add(Conv2D(32, kernel_size=(3, 3), activation='relu', input_shape=(48,48,1)))
model.add(Conv2D(64, kernel_size=(3, 3), activation='relu'))
model.add(MaxPooling2D(pool_size=(2, 2)))
model.add(Dropout(0.25))

model.add(Conv2D(128, kernel_size=(3, 3), activation='relu'))
model.add(MaxPooling2D(pool_size=(2, 2)))
model.add(Conv2D(128, kernel_size=(3, 3), activation='relu'))
model.add(MaxPooling2D(pool_size=(2, 2)))
model.add(Dropout(0.25))

model.add(Flatten())
model.add(Dense(1024, activation='relu'))
model.add(Dropout(0.5))
model.add(Dense(7, activation='softmax'))
```

```
from keras.callbacks import ModelCheckpoint
```

ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.776660

```
checkpoint = ModelCheckpoint('model_weights.h5', monitor='val_acc', verbose=1,
save_best_only=True, mode='max')
callbacks_list = [checkpoint]
num_train = 28821
num_val = 7066
batch_size = 64
num_epoch = 50
model.compile(loss='categorical_crossentropy',optimizer=Adam(lr=0.0001, decay=1e-
6),metrics=['accuracy'])
history = model.fit_generator(generator=train_generator,
                                steps_per_epoch=train_generator.n//train_generator.batch_size,
                                epochs=num_epoch,
                                validation_data = validation_generator,
                                validation_steps = validation_generator.n//validation_generator.batch_size,
                                callbacks=callbacks_list
                                )
model.save_weights('model.h5')
```

```
model_json = model.to_json()
with open('model.json', 'w') as json_file:
    json_file.write(model_json)
```

```
# plot the evolution of Loss and Acuracy on the train and validation sets

import matplotlib.pyplot as plt

plt.figure(figsize=(20,10))
plt.subplot(1, 2, 1)
plt.suptitle('Optimizer : Adam', fontsize=10)
plt.ylabel('Loss', fontsize=16)
plt.plot(history.history['loss'], label='Training Loss')
plt.plot(history.history['val_loss'], label='Validation Loss')
plt.legend(loc='upper right')

plt.subplot(1, 2, 2)
plt.ylabel('Accuracy', fontsize=16)
plt.plot(history.history['accuracy'], label='Training Accuracy')
plt.plot(history.history['val_accuracy'], label='Validation Accuracy')
plt.legend(loc='lower right')
plt.show()
```

```
# show the confusion matrix of our predictions

# compute predictions
predictions = model.predict_generator(generator=validation_generator)
y_pred = [np.argmax(probas) for probas in predictions]
y_test = validation_generator.classes
class_names = validation_generator.class_indices.keys()
```

Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.776660

```
from sklearn.metrics import confusion_matrix
import itertools

def plot_confusion_matrix(cm, classes, title='Confusion matrix', cmap=plt.cm.Blues):
    cm = cm.astype('float') / cm.sum(axis=1)[:, np.newaxis]
    plt.figure(figsize=(10,10))
    plt.imshow(cm, interpolation='nearest', cmap=cmap)
    plt.title(title)
    plt.colorbar()
    tick_marks = np.arange(len(classes))
    plt.xticks(tick_marks, classes, rotation=45)
    plt.yticks(tick_marks, classes)

    fmt = '.2f'
    thresh = cm.max() / 2.
    for i, j in itertools.product(range(cm.shape[0]), range(cm.shape[1])):
        plt.text(j, i, format(cm[i, j], fmt),
                 horizontalalignment=‘center’,
                 color=‘white’ if cm[i, j] > thresh else ‘black’)

    plt.ylabel('True label')
    plt.xlabel('Predicted label')
    plt.tight_layout()

# compute confusion matrix
cnf_matrix = confusion_matrix(y_test, y_pred)
np.set_printoptions(precision=2)

# plot normalized confusion matrix
plt.figure()
plot_confusion_matrix(cnf_matrix, classes=class_names, title='Normalized confusion matrix')
plt.show()
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:4: UserWarning:
`Model.predict_generator` is deprecated and will be removed in a future version. Please use
`Model.predict`, which supports generators.
  after removing the cwd from sys.path.
<Figure size 432x288 with 0 Axes>
```

9. CONCLUSION

The research presented in this paper aimed to investigate the effectiveness of a machine learning model for emotion detection using data. The results obtained from our experiments indicated promising performance of the proposed model in accurately classifying emotions based on facial expressions. The model achieved an average accuracy of 97% on the test data, with precision, recall, F1-score,

ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.776660

specificity, and sensitivity values also showing favourable results. The accuracy graph depicted a clear upward trend with increasing epochs during the model training process. This indicates that the model's accuracy improved over time, which is an encouraging sign for the potential of the proposed approach in real-world applications. The accuracy metric is a crucial evaluation measure, as it reflects the overall correctness of the model's predictions. Furthermore, the confusion matrix provided insights into the model's performance for each emotion class. The model showed relatively high precision and recall values for most of the emotion classes, indicating that it was able to correctly identify emotions with minimal false positives and false negatives. However, the model struggled slightly with the "Surprise" emotion, which had a lower recall value, indicating that there is room for improvement in accurately detecting this particular emotion.

The results obtained from our experiments have several implications for future research in the field of emotion detection. Firstly, our findings highlight the potential of machine learning models in accurately detecting emotions from facial expressions. This can have various applications, such as in human-computer interaction, affective computing, and psychological research. Further research can explore more advanced machine learning algorithms or deep learning techniques to improve the accuracy and robustness of emotion detection models. Additionally, our research also has some limitations that should be considered. The dataset used in this research was relatively small, and this could impact the model's performance. Larger datasets with a more balanced distribution of emotions can be used to further validate the proposed model.

In conclusion, our research demonstrates the potential of machine learning in accurately detecting emotions from facial expression. The proposed model achieved promising results in terms of accuracy, precision, recall, F1-score, specificity, and sensitivity. The findings of this research have implications for various applications such as human-computer interaction, affective computing, and psychological research. However, the study also has limitations that should be considered, which is the relatively small dataset. Future research can explore more advanced machine learning algorithms, deep learning techniques, and real-world datasets to improve the accuracy and generalizability of emotion detection models. Overall, this research contributes to the growing body of literature on emotion detection using machine learning and opens avenues for further investigation in this field.

10. REFERENCES

1. Ekman, R. What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS) (Oxford University Press, 1997).

Journal of
Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.776660

2. S. E. Kahou, V. Michalski, K. Konda, R. Memisevic, and C. Pal, 'Recurrent neural networks for emotion recognition in video.' Proceedings of the 2015 ACM on International Conference on Multimodal Interaction. 467-474. ACM.
3. Nwosu, L.,Wang, H., Lu, J., Unwala, I., Yang, X., Zhang, T. Deep convolutional neural network for facial expression recognition using facial parts. In 2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intel- ligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/ PiCom/DataCom/Cyber SciTech) 1318–1321 (IEEE, 2017).
4. Yang, B., Xiang, X., Xu, D., Wang, X. & Yang, X. 3d palm print recognition using shape index representation and fragile bits.
5. W. Mou, O. Celiktutan and H. Gunes, 'Group-level arousal and valence recognition in static images: Face, body and context.,' in 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Ljubljana, Slovenia, 2015, Multimed. Tools Appl. 76(14), 15357–15375 (2017).
6. L. Tan, K. Zhang, K. Wang, X. Zeng, X. Peng, and Y. Qiao, 'Group Emotion Recognition with Individual Facial Emotion CNNs and Global Image-based CNNs,' in 19th ACM International Conference on Multimodal Interaction, Glasgow, UK, 2017.
7. Kumar, N. & Bhargava, D. A scheme of features fusion for facial expression analysis: A facial action recognition. J. Stat. Manag. Syst. 20(4), 693–701 (2017).
8. Tzimiropoulos, G. & Pantic, M. Fast algorithms for fitting active appearance models to unconstrained images. Int. J. Comput. Vis.122(1), 17–33 (2017).
9. M. Sajid, NI. Ratyal, N. Ali, B. Zafar, S. H. Dar, M. T. Mahmood, and Y. B Joo, 'The impact of asymmetric left and asymmetric right face images on accurate age estimation.' Math Probl Eng 2019: 1–10.
10. Zhao, G. & Pietikainen, M. Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Trans. Pattern Anal. Mach. Intell. 29(6), 915–928 (2007).
11. Ahmadinia, M. et al. Energy-efficient and multi-stage clustering algorithm in wireless sensor networks using cellular learning automata. IETE J. Res. 59(6), 774–782 (2013).
12. Zhao, X., Liang, X., Liu, L., Li, T., Han, Y., Vasconcelos, N., Yan, S. Peak-piloted deep network for facial expression recognition. In European Conference on Computer Vision 425–442 (Springer, 2016).
13. Zhang, H., Jolfaei, A. & Alazab, M. A face emotion recognition method using convolutional neural network and image edge computing. IEEE Access 7, 159081–159089 (2019).
14. Goodfellow, I. J., Erhan, D., Carrier, P. L., Courville, A., Mirza, M., Hamner, B., Cukierski, W., Tang, Y., Thaler, D., Lee, D.-H. et al. Challenges in representation learning: A report on three machine learning contests. In International Conference on Neural Information Processing 117–124 (Springer, 2013).
15. Yu, Z. & Zhang, C. Image based static facial expression recognition with multiple deep network learning. In Proceedings of the 2015 ACM on International Conference on Multimodal Interaction 435–442 (2015).
16. Niu, H. et al. Deep feature learnt by conventional deep neural network. Comput. Electr. Eng. 84, 106656 (2020).
17. Pantic, M., Valstar, M., Rademaker, R., & Maat, L. Web-based database for facial expression analysis. In 2005 IEEE International Conference on Multimedia and Expo 5 (IEEE, 2005).
18. Wang, X., Feng, X., & Peng, J. A novel facial expression database construction method based on web images. In Proceedings of the Third International Conference on Internet Multimedia Computing and Service 124–127 (2011).
19. Mayer, C., Eggers, M. & Radig, B. Cross-database evaluation for facial expression recognition. Pattern Recognit. Image Anal. 24(1), 124–132 (2014).
20. Tang, Y. Deep learning using linear support vector machines. arXiv preprint arXiv:1306.0239 (2013).
21. Gan, Y. Facial expression recognition using convolutional neural network. In Proceedings of the 2nd International Conference on Vision, Image and Signal Processing 1–5 (2018).

ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.776660

22. Li, C.-E. J. & Zhao, L. Emotion recognition using convolutional neural networks. In Purdue Undergraduate Research Conference 63 (2019).
23. Lv, Y., Feng, Z., & Xu, C. Facial expression recognition via deep learning. In 2014 International Conference on Smart Computing 303–308 (IEEE, 2014).
24. Mollahosseini, A., Chan, D., & Mahoor, M. H. Going deeper in facial expression recognition using deep neural networks. In 2016 IEEE Winter Conference on Applications of Computer Vision (WACV) 1–10 (IEEE, 2016).
25. Hahnloser, R. H., Sarpeshkar, R., Mahowald, M. A., Douglas, R. J. & Seung, H. S. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. Nature 405(6789), 947–951 (2000).
26. Patil, M. N., Iyer, B., & Arya, R. Performance evaluation of PCA and ICA algorithm for facial expression recognition application. In Proceedings of Fifth International Conference on Soft Computing for Problem Solving 965–976 (Springer, 2016).
27. Christou, N. & Kanojiya, N. Human facial expression recognition with convolution neural networks. In Third International Congress on Information and Communication Technology 539–545 (Springer, 2019).
28. Niu, B., Gao, Z. & Guo, B. Facial expression recognition with LBP and ORB features. Comput. Intell. Neurosci. 2021, 1–10 (2021).
29. González-Lozoya, S. M. et al. Recognition of facial expressions based on cnn features. Multimed. Tools Appl. 79, 1–21 (2020).
30. Christy, A., Vaithyasubramanian, S., Jesudoss, A. & Praveena, M. A. Multimodal speech emotion recognition and classification using convolutional neural network techniques. Int. J. Speech Technol. 23, 381–388 (2020).
31. Niu, H. et al. Deep feature learnt by conventional deep neural network. Comput. Electr. Eng. 84, 106656 (2020).
32. Wang, F. et al. Emotion recognition with convolutional neural network and EEG-based EFDMs. Neuropsychologia 1(146), 107506 (2020).
33. Wang, F. et al. Emotion recognition with convolutional neural network and eeg-based efdms. Neuropsychologia 146, 107506 (2020).
34. Canedo, D. & Neves, A. J. Facial expression recognition using computer vision: A systematic review. Appl. Sci. 9(21), 4678 (2019).
35. Nonis, F., Dagnes, N., Marcolin, F. & Vezzetti, E. 3d approaches and challenges in facial expression recognition algorithms—A literature review. Appl. Sci. 9(18), 3904 (2019).
36. Hans, A. S. A. & Smitha, R. A CNN-LSTM based deep neural networks for facial emotion detection in videos. Int. J. Adv. Signal Image Sci. 7(1), 11–20 (2021).
37. Bodapati, J. D. & Veeranjaneyulu, N. Facial emotion recognition using deep cnn based features (2019).
38. Haddad, J., Lézoray, O., & Hamel, P. 3d-cnn for facial emotion recognition in videos. In International Symposium on Visual Com- puting 298–309 (Springer, 2020).
39. Hussain, S. A. & Ahlam, S. A. A. B. A real time face emotion classification and recognition using deep learning model. J. Phys. Conf. Ser. 1432(1), 012087 (2020).
40. Singh, S. & Nasoz, F. Facial expression recognition with convolutional neural networks. In 2020 10th Annual Computing and Communication Workshop and Conference (CCWC) 0324–0328 (IEEE, 2020).
41. Shan, C., Gong, S. & McOwan, P. W. Facial expression recognition based on local binary patterns: a comprehensive study. Image Vis. Comput. 27(6), 803–816 (2009).
42. https://www.kaggle.com/msambare/fer2013. In FER-2013|Kaggle. Accessed 20 Feb 2021.
43. https://github.com/Tanoy004/Facial-Own-images-for-test. Accessed 20 Feb 2021.
44. O. A. Montesinos López, A. Montesinos López, J. Crossa, 'Overfitting, Model Tuning, and Evaluation of Prediction Performance in Multivariate Statistical Machine Learning Methods for Genomic Prediction.' Springer 2022, Cham. https://doi.org/10.1007/978-3-030-89010-0_4